A Note on Averaging Rotations

W. Dan Curtis[1], Adam L. Janin[1], and Karel Zikan[1]

Abstract

Computing average rotations is a fundamental sensor-fusion problem. In Virtual Reality, the problem naturally arises in connection with position sensing. We derive two methods for additive averaging of rotations. One of these methods, based on quarternions, is limited to averages of 3-dimensional rotations, but the other method, based on singular value decomposition, works in any euclidean space.

Section 1 Introduction

Computing average rotations is a fundamental sensor-fusion problem. In Virtual Reality, the problem arises in connection with position sensing, for instance, when the orientation of a person (or an object) needs to be determined. All physical sensors are somewhat inaccurate and some are also subject to "noise". The inaccuracy of sensors results in incorrect alignment of images, the "noise" produces image jitter. The effects of both problems can be reduced by averaging data from different sensors, or by (weighted) averaging of data from a single sensor over time.

Let us consider two general rotations given by $n \times n$ matrices U_1 and U_2. It is not immediately obvious how to average them. We need to find a rotation matrix U that best resembles both U_1 and U_2. The matrix $\frac{1}{2}(U_1 + U_2)$ does not represent a rotation, nor does any multiple of this matrix. Averaging Euler's angles also does not work—we get different answers for different Cartesian coordinate systems.

We show in this paper that a sensible average of the two rotations is given by

$$U_a = XY, \text{ where } XDY = U_1 + U_2 \tag{1}$$

is a *singular value decomposition* (SVD) of $U_1 + U_2$. The matrix U_a is the projection of $U_1 + U_2$ onto the space of rotation matrices, that is, U_a maximizes the inner product

$$\langle U_1 + U_2, U \rangle = \text{trace}\big[(U_1 + U_2)U^T\big]. \tag{2}$$

[1] Boeing Computer Services, Research and Technology, P.O. Box 24346, M/S 7L-21, Seattle, WA 98124–0346

0-7803-1363-1/93 $3.00 © 1993 IEEE

More importantly, U_a solves the minimization problem for the average square penalty

$$\min_U \int_{||x||=1} \left[||(U_1 - U)x||^2 + ||(U_2 - U)x||^2\right] d(x) \tag{3}$$

for displacement between the U_1, U_2, and U rotations of the unit sphere. In other words, U_a minimizes (the square of) the *effects* of the difference between the rotations.

We also show that for 3-dimensional rotations, where rotations can be expressed in terms of quaternions, we can average rotations by adding and scaling quaternions. We show that if q_j is the quaternion corresponding to the rotation U_j (where $j = 1, 2$), then the optimization problems (2) and (3) are solved by the rotation corresponding to the quaternion q_a given by

$$q_a = \begin{cases} \frac{(q_1+q_2)}{\lambda}, & \text{if } \langle q_1, q_2 \rangle \geq 0; \\ \frac{(q_1-q_2)}{\mu}, & \text{otherwise}; \end{cases} \tag{4}$$

where $\lambda = ||q_1 + q_2||$ and $\mu = ||q_1 - q_2||$.

We thus have the choice of two computational procedures for averaging two rotations in three dimensions; we can either work directly with the rotation matrices, or we can work with the quaternions. The quaternion method may be preferred over the direct method on the computational grounds, particularly if the quaternions are readily available. The quaternion-based approach need be used with an utmost care, however. If we average three or more rotations, or if we seek a ***weighted average*** of two rotations, the quaternion approach can become extraordinarily complex. In general, the naive approach of simply averaging and normalizing the quaternions will not produce the desired answer. We therefore advise the exclusive use of the SVD-based approach, save in the simple two rotation case. An easy generalization of our derivation shows that if XDY is a singular value decomposition of $\sum w_j U_j$, then $U_a = XY$ gives the minimum of

$$\int_{||x||=1} \sum_j w_j ||(U_j - U)x||^2 d(x). \tag{5}$$

Section 2 The SVD Approach

We define the $n \times n$ rotation U_a to be the *average* of two other $n \times n$ rotations, U_1 and U_2, if and only if U_a solves the optimization problem (3). Using the uniform distribution of points on unit sphere, (3) is, up to a scaling, the definition of

$$E_x\left\{||(U_1 - U)x||^2 + ||(U_2 - U)x||^2\right\}. \tag{6}$$

The roots of the SVD approach to averaging rotations go back to the so-called *Orthogonal Procrustes Problem* of Green [2] and Shonemann [3]. Let $||\cdot||_F$ denote the Frobenius matrix norm, that is, $||A||_F^2 = \text{trace}(AA^T)$.

Lemma 2.1 (Green and Schonemann) *Given two $n \times k$ matrices A and B, the problem*

$$\min_U ||A - UB||_F \tag{7}$$

is solved when $U = XY$, where

$$XDY$$

is an SVD of AB^T.

Proof of this result can be found, for instance, in Golub and Van Loan [1].

We now state and prove the main result.

Theorem 2.2 *The average of the rotations U_1 and U_2 is the rotation $U_a = XY$, where XDY is a SVD of $U_1 + U_2$.*

Proof: Clearly, U_a solves (7) iff it also solves

$$\min_U ||A - UB||_F^2.$$

We note that $||A - UB||_F^2 = ||A||_F^2 + ||B||_F^2 - 2\langle A, UB\rangle$, where $\langle A, UB\rangle$ denotes the inner product of matrices, that is, trace(AB^TU^T). Ignoring the constant terms and rearranging the inner product in obvious ways, we see that (7) is also equivalent to

$$\max_U \left\langle AB^T, U\right\rangle. \tag{8}$$

Let us now consider the effects of the rotations U_1 and U_2 on any set of n orthonormal vectors v_j. These can be arranged into an orthogonal matrix V, and we can verify by a direct computation that

$$||(U_1V, U_2V) - (UV, UV)||_F^2 = \sum_{j=1}^{n} ||(U_1 - U)v_j||^2 + ||(U_2 - U)v_j||^2. \quad (9)$$

Since uniform sampling of orthogonal matrices V produces n sets of points $\{v_j\}$, each uniformly sampled from the unit sphere, (9) implies that

$$E_V\Big\{||(U_1V, U_2V) - U(V, V)||^2\Big\} = nE_x\Big\{||(U_1 - U)x||^2 + ||(U_2 - U)x||^2\Big\}. \quad (10)$$

The upshot of (10) is that it suffices to minimize $E_V\{||(U_1V, U_2V) - U(V, V)||^2\}$ to find U_a.

By setting $A = (U_1V, U_2V)$ and $B = (V, V)$ we see that, in this special case, (8) becomes

$$\max_U \langle U_1 + U_2, U\rangle, \quad (11)$$

since the term $VV^T = I$ can be dropped. By the Orthogonal Procrustes Problem, the maximum of (11) is achieved when $U_a = XY$. But note that (11) is independent of V! The expected value

$$E_V\{||(U_1V, U_2V) - U(V, V)||^2\}$$

is thus minimized by the rotation U_a. ***QED***

Section 3 The Quaternion Approach

The set of quaternions of unit length is just the 3–sphere in $\mathbf{R}^4$,

$$S^3 = \{q \in \mathbf{R}^4 | \, || \, q \, || = 1\}.$$

Using quaternion multiplication, S^3 becomes a Lie group and there is a two-fold covering projection

$$\rho \, : \, S^3 \longrightarrow SO(3)$$

given by

$$\rho(q)\mathbf{v} = \pi(qu\bar{q})$$

for all $\mathbf{v} \in \mathbf{R}^3$, where

$$u = \begin{bmatrix} 0 \\ \mathbf{v} \end{bmatrix}, \tag{12}$$

π projects onto the last three components of a quaternion, and $\bar{q} = (\alpha, -\mathbf{p})$ is the *conjugate* of the quaternion $q = (\alpha, \mathbf{p})$. (Note that the conjugate is the multiplicative inverse for $q \in S^3$.)

Before proving the next theorem, we consider two lemmas from linear algebra which will be useful in the proof of the theorem. For any matrix A, let S_A denote the sum of the singular values of A.

Lemma 3.1 *Let A be an $n \times n$ matrix. Then*

$$|\text{trace}(A)| \leq S_A$$

with equality if and only if A is symmetric and (positive or negative) semi-definite.
Proof: Construct an SVD for A

$$A = UDV^T,$$

where, as usual, U and V are orthogonal and $D = \text{diag}(\sigma_1, \ldots, \sigma_n)$ with the diagonal entries non-negative and in non-increasing order. Suppose $\sigma_i > 0$ for $1 \leq i \leq r$ and $\sigma_i = 0$ for $i > r$. If $u_1, \ldots, u_n$ and $v_1, \ldots, v_n$ are the column vectors of U and V respectively, then

$$A = \sum_{i=1}^{r} \sigma_i u_i v_i^T.$$

Then we have

$$\text{trace}(A) = \sum_{i=1}^{r} \sigma_i \text{trace}(u_i v_i^T) = \sum_{i=1}^{r} \sigma_i u_i^T v_i.$$

Therfore we have

$$|\text{trace}(A)| \leq \sum_{i=1}^{r} \sigma_i \parallel u_i \parallel\parallel v_i \parallel = \sum_{i=1}^{r} \sigma_i = S_A,$$

which proves the asserted inequality. We note from these calculations that trace(A) $= S_A$ exactly when $u_i^T v_i = \| u_i \| \| v_i \|$ for $i = 1, \ldots, r$. But this latter condition holds iff u_i and v_i are in the same direction for $i = 1, \ldots, r$ and since these are unit vectors (the matrices are orthogonal) we see $u_i = v_i$ for $i = 1, \ldots, r$. Therefore, we actually have

$$A = \sum_{i=1}^{r} \sigma_i u_i u_i^T,$$

so that A is a symmetric, positive semi-definite matrix. A similar argument shows that trace(A) $= -S_A$ iff $u_i = -v_i$ for $i = 1, \ldots, r$, in which case

$$A = -\sum_{i=1}^{r} \sigma_i u_i u_i^T,$$

so that A is symmetric, negative semi-definite. ***QED***

Lemma 3.2 *Let A be an $n \times n$ matrix. Then there is an orthogonal matrix Q such that the matrix AQ is symmetric, positive semi-definite. For such a Q, if Y is* any *orthogonal matrix, then*

$$\text{trace}(AY) \leq \text{trace}(AQ)$$

with equality iff AY is symmetric, positive semi-definite.
Proof: Obtain an SVD for A so that $A = UDV^T$; then let $Q = VU^T$. We have

$$AQ = UDU^T$$

which is symmetric, positive semi-definite. For an arbitrary orthogonal matrix Y, note that A and AY have the *same* singular values since $AY = UDV^TY$ gives an SVD for AY. Thus we have

$$\text{trace}(AY) \leq S_{AY} = S_A = \text{trace}(AQ).$$

Note that equality holds iff trace(AY) $= S_{AY}$ which, as shown in the proof of Lemma 3.1, is true iff AY is symmetric, positive semi-definite. ***QED***

Thus, given rotations U_1, U_2, and U_a, the condition that U_a maximize (2) is equivalent to the condition that the matrix $(U_1 + U_2)U_a^T$ is symmetric, positive semi-definite. We now prove

Theorem 3.3 *Let $q_1 \neq -q_2$ be unit quaternions with corresponding rotations U_1 and U_2. Then the unit quaternion*

$$q_a = \frac{q_1 + q_2}{||q_1 + q_2||}$$

corresponds to a rotation U_a maximizing (2) if and only if

$$\langle q_1, q_2 \rangle \geq 0.$$

Proof: Let q be any unit quaternion, with corresponding rotation U. The mapping ρ defined above, is a group homomorphism, so the rotations corresponding to $q_1 q$, $q_2 q$, and $q_a q$ are, respectively, $U_1 U$, $U_2 U$, and $U_a U$. But

$$(U_1 U + U_2 U)(U_a U)^T = (U_1 + U_2)U_a^T,$$

so q_a corresponds to the average of U_1 and U_2 if and only if $q_a q$ corresponds to the average of $U_1 U$ and $U_2 U$. If we choose $q = \bar{q}_1$ we see it is sufficient to prove the theorem in the case $q_1 = 1$ and $U_1 = I$. This simplifies the calculations below. Thus we have $\rho(q_2) = U_2$ and $\rho(q_a) = U_a$, where

$$q_a = \frac{1 + q_2}{\lambda}, \quad \lambda = ||q_1 + q_2||,$$

and we will show:

(i) The matrix $A = (I + U_2)U_a^T$ is symmetric;
(ii) A is positive semi-definite if and only if $\langle 1, q_2 \rangle \geq 0$.

To show (i) let $v \in \mathbf{R}^3$ and let $u = (0, \mathrm{v})^T$ be the corresponding quaternion. Then

$$\begin{aligned} A\mathrm{v} &= U_a^T \mathrm{v} + U_2 U_a^T \mathrm{v} \\ &= \pi(\bar{q}_a u q_a + q_2 \bar{q}_a u q_a \bar{q}_2) \\ &= \frac{1}{\lambda^2}\pi((1 + \bar{q}_2)u(1 + q_2) + q_2(1 + \bar{q}_2)u(1 + q_2)\bar{q}_2) \\ &= \frac{1}{\lambda^2}\pi((1 + \bar{q}_2)u(1 + q_2) + (q_2 + 1)u(\bar{q}_2 + 1)) \end{aligned} \tag{13}$$

while

$$A^T \mathrm{v} = \frac{1}{\lambda^2}\pi((1 + q_2)u(1 + \bar{q}_2) + (\bar{q}_2 + 1)u(q_2 + 1)). \tag{14}$$

so that symmetry holds.

To prove (ii) let $q_2 = (\alpha, \mathbf{p})^T$ and compute $\langle A\mathbf{v}, \mathbf{v}\rangle$ for an arbitrary $\mathbf{v} \in \mathbf{R}^3$. Here we ignore the positive factor λ^{-2} which appears in the above equations. Thus, using notation as above, we show that

$$\langle \pi((1+q_2)u(1+\bar{q}_2) + (\bar{q}_2+1)u(q_2+1)), \mathbf{v}\rangle \geq 0, \quad \text{for all } v \in \mathbf{R}^3 \tag{15}$$

if and only if

$$\alpha = \langle 1, q_2 \rangle \geq 0.$$

After some quaternion algebra (and using the assumption that q_2 is a unit quaternion so that $\alpha^2 + |\mathbf{p}|^2 = 1$), the inequality (15) becomes

$$(\mathbf{p}\cdot\mathbf{v})^2 + (\alpha + \alpha^2)\mathbf{v}\cdot\mathbf{v} \geq 0. \tag{16}$$

This holds for all $\mathbf{v} \in \mathbf{R}^3$ if and only if $\alpha + \alpha^2 \geq 0$ and (since $\alpha^2 + |\mathbf{p}|^2 = 1$) $-1 \leq \alpha \leq 1$. These conditions hold iff $0 \leq \alpha \leq 1$ or $\alpha = -1$. But if $\alpha = -1$ then $\mathbf{p} = 0$ so that $q_2 = -1$, contrary to our assumption that $q_1 \neq -q_2$. Thus we conclude that (15) holds iff $0 \leq \alpha \leq 1$, as asserted. ***QED***

Section 4 Conclusions and Future Work

We have given a definition of the average of two rotations and shown how the singular value decomposition can be used to compute it. Since quaternions are often used to handle rotation problems in 3–space, we show that the average of two rotations corresponds to the (normalized) arithmetic mean of the unit quaternions associated to the two rotations, provided that the inner product of the quaternions is non-negative. The last condition is required because of the fact that the quaternion associated to a given rotation is determined only up to sign. Thus, if we have two quaternions representing two rotations and if the inner product of the quaternions is negative, we must reverse the sign on one of them before averaging.

Other matters remain to be settled: computing averages of more than two rotations and computing various weighted averages may be important for applications. Computing a weighted average of two rotations can be done using the SVD technique, as well as via quaternions. The two methods do not agree, however, and we would like to better understand the differences. Computing averages of more than two rotations can be done via the SVD method. The conditions for obtaining the corresponding averages in terms of quaternions may be more complicated—should we try to make all dot products of pairs of the quaternions be positive? (This is not always possible.)

Bibliography

[1] G. H. Golub and C. F. Van Loan. *Matrix Computations*. Johns Hopkins University Press, 1983.

[2] B. Green. The orthogonal approximation of an oblique structure in factor analysis. *Psychometrika 17*, 1952.

[3] P. Schonemann. A generalized solution of the orthogonal procrustes problem. *Psychometrika 31*, 1966.